

Multi-scale 3D-convolutional neural network for hyperspectral image classification

Murali Kanthi¹, Thogarcheti Hitendra Sarma², Chigarapalle Shoba Bindu¹

¹Department of Computer Science and Engineering, College of Engineering, Jawaharlal Nehru Technological University (JNTU), Anantapur, India

²Department of Information Technology, Vasavi College of Engineering, Hyderabad, India

Article Info

Article history:

Received Jul 29, 2021

Revised Nov 16, 2021

Accepted Nov 30, 2021

Keywords:

Convolutional neural network
Deep learning
Hyperspectral image
Multi-scale
Spatial and spectral information

ABSTRACT

Deep learning methods are state-of-the-art approaches for pixel-based hyperspectral images (HSI) classification. High classification accuracy has been achieved by extracting deep features from both spatial-spectral channels. However, the efficiency of such spatial-spectral approaches depends on the spatial dimension of each patch and there is no theoretically valid approach to find the optimum spatial dimension to be considered. It is more valid to extract spatial features by considering varying neighborhood scales in spatial dimensions. In this regard, this article proposes a deep convolutional neural network (CNN) model wherein three different multi-scale spatial-spectral patches are used to extract the features in both the spatial and spectral channels. In order to extract these potential features, the proposed deep learning architecture takes three patches various scales in spatial dimension. 3D convolution is performed on each selected patch and the process runs through entire image. The proposed is named as multi-scale three-dimensional convolutional neural network (MS-3DCNN). The efficiency of the proposed model is being verified through the experimental studies on three publicly available benchmark datasets including Pavia University, Indian Pines, and Salinas. It is empirically proved that the classification accuracy of the proposed model is improved when compared with the remaining state-of-the-art methods.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Murali Kanthi

Department of Computer Science & Engineering, College of Engineering

Jawaharlal Nehru Technological University (JNTU)

Anantapur, Andhra Pradesh, India

Email: murali.kanthi@gmail.com

1. INTRODUCTION

Hyperspectral image (HSI) analysis has become an active research area because of the reliability of the results in a wide range of earth monitoring applications viz, precision agriculture, geological mapping, environmental and climate observations, disaster management, health care, defense, and many others [1], [2]. Hyperspectral remote sensing collects large amount of data in the form of HSI, which are useful in a variety of applications [1]. Many supervised classifiers for HSI classification have been proposed in the literature [3]. Deep convolutional neural network (CNN) models have proved to be effective in extracting features resulting in improved classification accuracy of HSI [3], [4]. Extracting discriminant spatial-spectral features is the key factor to achieve high classification accuracy [5]-[7]. Various methods have been presented for extracting spatial-spectral features. A model presented by Chen *et al.* [8] on the concept of joint spatial-spectral classification, in which each pixel's spatial features are chosen and combined with spectral characteristics.

When limited training data is available, Yang *et al.* [9] applied transfer learning to improve classification performance by using deep CNN model with a two-branch network to extract combined spatial- spectral features. Chen *et al.* [10] 3D image patches defined via a spatial window have been used to extract spatial spectral information.

Hamida *et al.* [11] proposed and evaluated set of 3-D schemes that enable a joint spatial-spectral information processing by combining the traditional 3-D convolution operations to analyze series of volumetric representations of the HSI. Raviteja *et al.* [12] introduced hierarchical image fusion model for HIS segmentation to create image groups for merging the selected spectral features. He *et al.* [13] proposed multi-scale convolutional neural network (MS-CNN) to address the problem of low interclass and large intraclass variance by extracting deep multi-scale features from the hyperspectral image. Wan *et al.* [14] proposed multiscale graph convolutional network to operate convolution on irregular image regions for HSI classification. Meng *et al.* [15] developed a completely dense multiscale fusion network for HSI classification by providing feed-forward shortcut connections across the layers to retrieve hierarchical information from all the convolutional layers. Roy *et al.* [16] proposed hybrid spectral network (HybridSN) model, which combines spectral and spatial three-dimensional convolutional neural network (3D-CNN) with spatial two-dimensional convolutional neural network (2DCNN). It provides good classification accuracy with a small training sample. In similar manner, Kanthi *et al.* [17] introduced a 3D-CNN approach for HSI classification, that divides HSI data into 3D patches and extracts deep spectral and spatial information. This model produced relatively high classification accuracy. Another successful approach for classification of HSI is to use ensemble techniques and some techniques have been proposed to extract features by varying the spatial dimension of the pixel patch using different CNN models and then combining all the extracted features to perform classification [18]-[20]. For HSI classification, a multi-scale three-dimensional convolutional neural network (M3DCNN) has been proposed that extracts multi-scale spatial features and spectral features from HSI [21]. Mohan and Venkatesan [22] presented a hybrid convolutional neural network (HybridCNN) model based on multi-scale spatial-spectral information of HSI for classification. Initially, kernel principal component analysis (KPCA) is used for dimensionality reduction in the preprocessing then 3D-CNN is applied with different window sizes to extract the spectral-spatial features. Safari *et al.* [23] proposed a model in which several CNNs are merged to learn spatial-spectral characteristics at numerous scales. Han *et al.* [24] proposed a different scale two-steem CNN for multi-scale image classification. Recently, Sun *et al.* [25] For HSI data classification with spectral-spatial fusion, a localized spectral features and multi-scale spatial features convolution (LSMSC) network was developed for multi-scale spatial feature extraction and dimensionality. The training parameters used in the model are much more than a traditional 3×3 convolution. This model was tested on benchmark data sets with huge number of training samples. However, it decreases its generalization ability when fewer training samples available. Gong *et al.* [26] proposed multiscale squeeze-and-excitation pyramid pooling network (MSPN) model to overcome the “small sample problem” with multiscale 3D-CNN module, squeezing and excitation block, and pyramid pooling. However, the model is more complex as different modules used and the model can be enhanced to improve the performance.

The proposed article presents a multi-scale 3DCNN learning model, called MS-3DCNN, for pixel-based classification in hyperspectral images. In this proposed method, various spatial contexts of a specific pixel are analyzed to provide multi-scale 3D patches for the model for extracting spatial-spectral features from HSI cube. This key contributions of the current work are:

- Spatial-spectral approaches depends on the spatial dimension of each patch and there is no theoretically valid approach to find the optimum spatial dimension to be considered. To avoid this issue, spatial and spectral features are extracted using multiple spatial contexts in three layers simultaneously and all are fused for further classification.
- The proposed deep CNN model is tested in new Indian hyperspectral images and compared with the state-of-the-art approaches to empirically establish superior performance with fewer training examples.

The rest of this article is organized in the following manner. The proposed model description is provided in section 2. Details of experimental setup and data descriptions are presented in section 3. Finally, section 4 presents the conclusions and future scope of the proposed work.

2. RESEARCH METHODOLOGY

Introduction to the proposed work's motivation is provided in this section. Using an architectural diagram, this section describes in detail how the proposed model, which is referred to as the multi-scale three-dimensional convolutional neural network (MS-3DCNN) model. It is able to retrieve individual pixels in the form of multi-scale 3D patches in three different spatial contexts.

2.1. Motivation

Spatial-spectral approaches depends on the spatial dimension of each patch and there is no theoretically valid approach to find the optimum spatial dimension to be considered. To avoid this issue, in this present work, the spatial and spectral features are extracted using multiple spatial contexts in three layers simultaneously and all are fused for further classification. It can ben generalized to have more than three layers. Current study is confined to analyse the efficiency of a simple multi-scale 3DCNN (MS-3DCNN) with three layers.

2.2. The proposed model

The proposed multi-scale 3DCNN (MS-3DCNN) approach is described in detail in this section. As shown in Figure 1, the proposed model MS-3DCNN takes multi-scale 3D patches as input to obtain fused spatial-spectral deep features from the given HSI. Let the given hyperspectral image is represented as three-dimensional cube with dimensions $W \times W \times B$, where W, H shows image's spatial width and height, and number of spectral bands is denoted by B . As similar to the existing most popular CNN architectures [11], [16], [21], [22], [25], initially the number of bands are reduced using principal component analysis (PCA) and the number of spectral bands selected as 30 in case of Indian Pines (IP) dataset and 15 spectral bands for the remaining datasets used in the experimental study of this article Indian Institute of Space Science and Technology (IIST), Ahmedabad1 (AH1), and Ahmedabad2 (AH2) respectively.

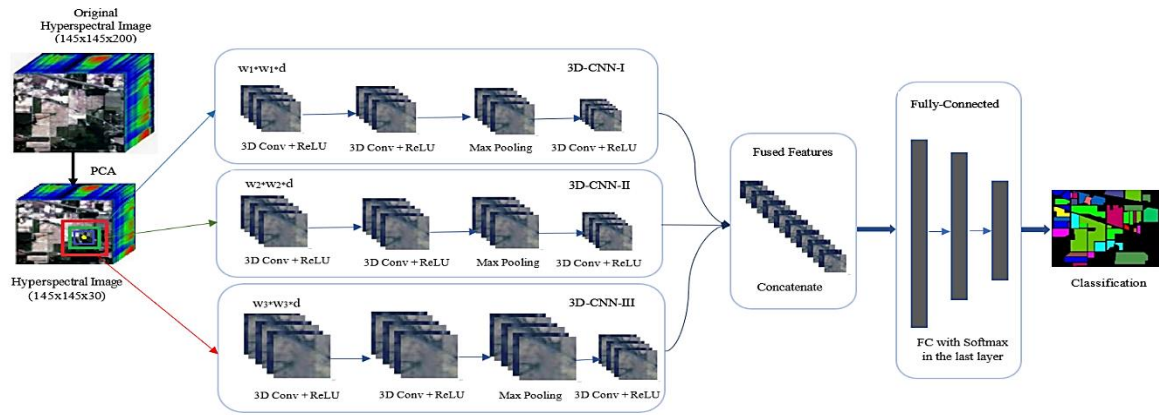


Figure 1. Overview of the proposed multi-scale 3DCNN (MS-3DCNN) model

In the MS-3DCNN model, spatial-spectral characteristics for individual pixel are retrieved in the form of multi-scale 3D patches in three separate spatial contexts. These multiscale 3D patches are fed to three 3D CNN models as shown in Figure 1. Each patch is of size $w_i \times w_i \times d$, where w_i is width and height of patch i , d is depth of patch and is number of patches of a particular pixel. In the current work, all experiments are conducted on the Google Colab Pro graphical processing unit (GPU) with 25.51 GB of RAM. Based on this configuration, we have chosen optimized three patches with sizes of $w_1 \times w_1 \times d = 13 \times 13 \times 30$, $w_2 \times w_2 \times d = 11 \times 11 \times 30$ and $w_3 \times w_3 \times d = 9 \times 9 \times 30$. As shown in Figure 1, in each layer, 3D CNN model contains three convolution layers (C_1, C_2 and C_3), max pooling layer (P) and three sets of filters $K_1 = 16, K_2 = 32$ and $K_3 = 64$ having sizes $3 \times 3 \times 7, 3 \times 3 \times 5$ and $3 \times 3 \times 3$, respectively. The max-pooling and batch normalization (BN) layers come after the first two convolutional layers, while the BN layer comes after last convolutional layer. The ReLU activation function is applied after every convolutional layer and max pooling with strides of $2 \times 2 \times 2$, as in (1).

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \tag{1}$$

For classification, the extracted features from various levels of spatial contexts are reshaped, concatenated, and sent to the fully connected layers fc_1, fc_2 and fc_3 . The dropout layer has been applied with the rate of 0.4% after every fully connected layer as a regularization mechanism to avoid the overfitting problem when there is limited availability of training samples. The activation function of each neuron in each of the fully connected layers are computed, as in (2).

$$Act_i(fc) = g(w_i(fc) * act_{i-1}(fc) + b_i) \tag{2}$$

where, $w_i(fc)$ is the weighted sum of the preceding layer's inputs and b_i is the bias. The ReLU activation function is represented by $g(\cdot)$. Finally, a soft-max probabilistic model is used to classify the data. Let $L=[L_i]_i$ be a number between 1 and n , and implies learned features after the entire model has been applied, as in (3).

$$Smax(L)_i = \frac{e^{L_i}}{\sum_{j=1}^k e^{L_j}} \text{ for } i = 1, 2, 3, \dots, n \quad (3)$$

For the HSI data, this is the Softmax function model. Finally, the argmax (arguments of maximum) function could be used to predict label, as in (4).

$$Class(X_i) = argmax\{Smax(L)_i\} \quad (4)$$

3. RESULTS AND DISCUSSION

3.1. Datasets

To assess the performance of proposed MS-3DCNN model, an experimental study was conducted on HSI datasets, including Pavia University (PU), Indian Pines (IP), and Salinas (SA). The first dataset, Indian Pines (IP), an airborne visible and infrared imaging spectrometer (AVIRIS) sensor captured the scene of Indian pines test site, which have 220 spectral bands and 145×145 pixels. The second dataset, Pavia University (PU) scene, a reflective optics system imaging spectrometer (ROSIS) sensor captured the scene of Pavia University at the time-of-flight campaign over Pavia. It has a resolution of 610×340 pixels and a total of 103 spectral bands. The third dataset, An AVIRIS sensor captured the Salinas (SA) landscape over Salinas Valley, California. There are 512×217 samples and 204 spectral bands in all. With 16, 9, and 16 class types, ground truths are provided for IP, PU, and SA, respectively. In addition, three new datasets IIST, AH1, and AH2 have been used to check the efficiency of the proposed model. These datasets are collected by the Indian space research organization (ISRO) with airborne visible and infrared imaging spectrometer-next generation (AVIRIS-NG) sensor, India [27]. The IIST dataset has 202×153 samples and 138 spectral bands, with 6 classes in the ground truth. The AH1 dataset has 351 spectral bands with a size of 300×200 pixels and the ground truth contains 5 classes. The AH2 dataset has 370 spectral bands with a size of 300×200 pixels and the ground truth contains 7 classes.

3.2. Experimental setup

The efficiency of the proposed model MS-3DCNN is analyzed by taking randomly selecting 20% of examples as training set, 10% selected as validation set and 70% as test set from each dataset. The Adam optimizer is used in the optimization process, along with a categorical cross-entropy with learning rate 0.001 and decay $1e-06$. The model has been trained for 100 epochs with batch normalization size 32. The experiments are repeated 10 times on each data set and the average results are presented. The conventional assessment measures: average accuracy (AA), overall accuracy (OA), and kappa (K) coefficient are used for comparing various similar models. The generalisation ability of the proposed model was tested by using 70% of the data from each dataset as a test set once the model was created.

3.3. Models for comparative study

The presented MS-3DCNN model's results are compared to the of other recent deep CNN models for HSI classification. Including 3DCNN [11], M3DCNN [21], HybridSN [16], HybridCNN [22], and LSMSC [25]. Table 1 provides the classification accuracy obtained by all these methods and it shows that the proposed model's classification accuracy is better than that of different approaches on the benchmark datasets in terms of evaluation metrics such as OA, AA, and kappa.

Table 1. Classification accuracies (in %) on Indian Pines, Pavia University, and Salinas datasets

Model	Indian Pines			Pavia University			Salinas		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
3D-CNN [11]	91.14	91.59	89.99	96.54	98.12	95.53	93.98	97.07	93.38
M3D-CNN [21]	95.33	94.72	96.48	95.78	94.52	96.10	94.99	94.22	96.31
HybridSN [16]	99.22	98.56	99.12	99.93	99.83	99.91	99.99	99.99	99.99
HybridCNN [22]	99.80	99.72	99.76	99.99	99.98	99.99	100	100	100
LSMSC [25]	96.71	98.08	96.11	99.22	99.25	98.95	98.70	99.35	98.54
Proposed Method	99.89	98.87	99.24	99.99	99.97	99.99	100	100	100

The accuracies for 3DCNN [11], M3DCNN [21], HybridSN [16], HybridCNN [22], and LSMSC [25] methods are taken from their respective papers, and the results are computed using publicly available code for the comparative methods. When compared to the 3DCNN [11] and the M3DCNN [21], the proposed model's overall-accuracy, average-accuracy, and kappa values are significantly better. Compared with the HybridSN [16] and the HybridCNN [22], the proposed method produces slightly better results in almost all cases. Moreover, the proposed method produced an improvement on overall accuracy when compared to LSMSC in the range 1-3%. It is worth mentioning that in the HybridCNN and the LSMSC models, 20% of samples randomly chosen for training process. The HybridSN model used 30% samples for training. In the proposed method 20% samples used from each class to train the model but still achieves better accuracy.

Table 2 shows the efficacy of proposed approach according to size of the training data. With fewer training data, the presented model achieved a higher classification accuracy than LSMSC. Experiments on the other three datasets, IIST, AH1, and AH2, are conducted to verify the efficiency and robustness of the model. We have compared our method with 3D-CNN and HybridSN models using their publicly available code. Other methods could not be compared as their code is not available. Table 3 shows that the presented model achieved 2 to 3% improvement on the new datasets.

Table 2. Classification accuracies of the proposed method by reducing the amount of training samples

Dataset	5 % Training data			10% Training data		
	OA	AA	Kappa	OA	AA	Kappa
IP	96.54	95.72	96.54	99.01	98.03	98.98
PU	99.62	98.71	99.67	99.89	99.36	99.73
SA	99.79	99.64	99.72	99.98	99.97	99.93

Table 3. Classification accuracies (in %) on the IIST, AH1, and AH2 datasets

Model	IIST dataset			AH1 dataset			AH2 dataset		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
3D-CNN [11]	94.26	91.99	90.48	80.99	82.13	78.17	70.06	69.30	67.93
HybridSN [16]	93.74	89.20	92.16	85.69	85.03	83.79	79.55	76.71	75.82
Proposed Method	96.42	91.51	94.13	87.24	87.15	85.74	80.10	77.05	76.72

Table 4 illustrates the training time and testing time for the state-of-the-art methods and proposed model on SA, PU, and IP datasets. The training time is counted in minutes, while testing time is counted in seconds. When compared to HybridCNN, the model training time is less, but when compared to 3DCNN and HybridSN, the training time is higher. Since the model uses more test data, it takes significantly more to test than HybridCNN. Table 5 shows the training time and testing time for the 3DCNN, HybridSN and proposed model on IIST, AH1, and AH2 datasets. On the new datasets, the suggested model's train and test times are slightly longer, despite the fact that it achieves greater classification accuracies.

Table 4. Training time(min) and testing time(sec) for SA, PU, IP datasets

Model	SA dataset		PU dataset		IP dataset	
	Training time	Test time	Training time	Test time	Training time	Test time
3D-CNN [11]	62	78	52	65	45	52
HybridSN [16]	50	64	45	60	40	50
HybridCNN [22]	122	27	112	23	74	11
Proposed Method	80	82	76	68	52	58

Table 5. Training time(min) and testing time(sec) for IIST, AH1, AH2 datasets

Model	IIST dataset		AH1 dataset		AH2 dataset	
	Training time	Test time	Training time	Test time	Training time	Test time
3D-CNN [11]	123	180	91	112	92	110
HybridSN [16]	120	176	88	104	86	102
Proposed Method	128	192	93	115	94	113

Table 6, Table 7, and Table 8, respectively, illustrate the proposed model's accuracy of each class as well as Precision, Recall, and F1- Score for the IIST, AH1, and AH2 datasets. The impact of spatial patch sizes over the performance of the proposed model is shown in Table 9. Similar analysis is done in [16] and it is concluded that the efficiency of the model drops if the spatial patch size is increased and it is computationally infeasible. It is also observed that fusion of features extracted with few small patches using different spatial window sizes can improve performance of the model.

Table 6. IIST dataset: number of samples, overall accuracy, precision, recall and F1-score

Class	Samples	OA	Precision	Recall	F1-Score
Class1	2288	96.58	0.95	0.97	0.96
Class2	3218	88.65	0.87	0.88	0.87
Class3	1735	95.90	0.94	0.96	0.94
Class4	18540	94.63	0.94	0.95	0.95
Class5	8059	93.27	0.93	0.94	0.92
Class6	2531	90.73	0.89	0.90	0.87

Table 7. AH1 dataset: number of samples, overall accuracy, precision, recall and F1-score

Class	Samples	OA	Precision	Recall	F1-Score
Class1	4210	90.70	0.88	0.91	0.88
Class2	9711	91.95	0.92	0.92	0.92
Class3	15318	82.41	0.84	0.85	0.85
Class4	21079	85.95	0.87	0.86	0.87
Class5	9682	97.71	0.99	0.98	0.98

Table 8. AH2 dataset: Number of samples, overall accuracy, precision, recall and F1-score

Class	Samples	OA	Precision	Recall	F1-Score
Class1	10741	75.46	0.75	0.73	0.75
Class2	11196	73.15	0.74	0.72	0.71
Class3	7901	89.04	0.89	0.87	0.88
Class4	14595	82.35	0.79	0.79	0.75
Class5	3264	85.14	0.88	0.82	0.86
Class6	6971	82.37	0.79	0.76	0.77
Class7	5332	79.84	0.78	0.75	0.74

Table 9. Impact of the spatial patch size on the performance (in %) of proposed model

Dataset	Spatial patch sizes		
	$w_1=13 \times 13, w_2=11 \times 11, w_3=9 \times 9$	$w_1=11 \times 11, w_2=9 \times 9, w_3=7 \times 7$	$w_1=9 \times 9, w_2=7 \times 7, w_3=5 \times 5$
SA	100	99.21	97.86
PU	99.99	98.93	96.87
IP	99.89	98.82	97.10
IIST	96.42	95.38	94.85
AH1	87.24	86.48	85.76
AH2	80.10	78.97	77.85

Hence, in the proposed work, three patches are being used by gradually increasing the spatial dimensions of the kernel for convolution. Due to the limitation of computing environment, Google Colab Pro GPU with 25.51 GB of random access memory (RAM), in this article, the studies are reported with three patch sizes of $w_1 \times w_1 \times d = 13 \times 13 \times 30$, $w_2 \times w_2 \times d = 11 \times 11 \times 30$ and $w_3 \times w_3 \times d = 9 \times 9 \times 30$ as inputs to the proposed model. The classification map of the proposed model of Indian Pines (IP) dataset is compared with its ground truth. It is indicating that the percentage of misclassification is quite minimal. Figure 2(a) represents the ground truth image of IP dataset, Figure 2(b) represents the model classification map of IP dataset, and Figure 2(c) represents reflective class legends of the image of IP dataset respectively.

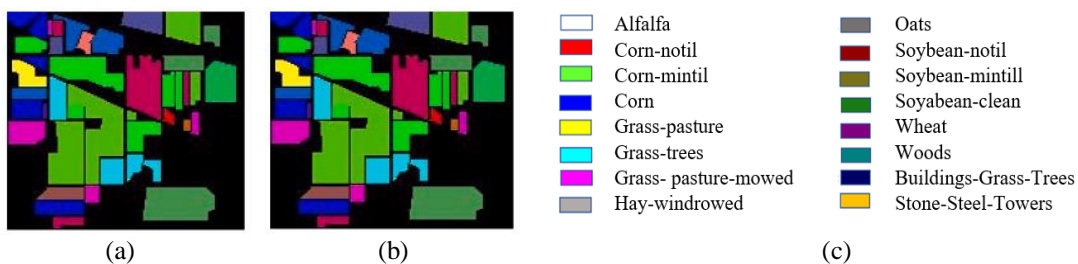


Figure 2. IP dataset; (a) ground-truth-image, (b) classification map, and (c) reflective class legends

The proposed model's classification map of Pavia University (PU) is compared with its ground truth. The classification map is showing that the percentage of misclassification is quite minimal. Figure 3(a) represents the ground truth image of PU dataset. Figure 3(b) represents the model classification map of PU dataset, and Figure 3(c) represents reflective class legends of the image of Pavia University dataset respectively.

The classification map of the proposed model of Salina (SA) dataset is compared with its corresponding ground truth. The classification map demonstrates that the percentage of misclassifications is extremely low. Figure 4(a) represents the ground truth image of SA dataset, Figure 4(b) represents the model classification map of SA dataset, and Figure 4(c) represents reflective class legends of the image of SA dataset respectively.

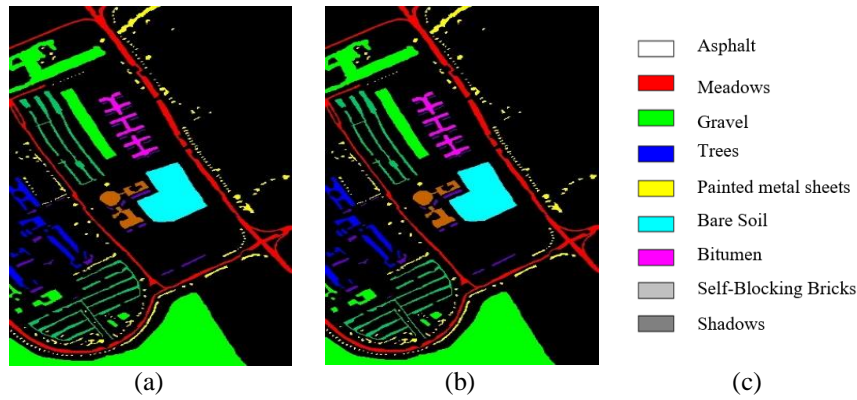


Figure 3. PU dataset; (a) ground-truth-image, (b) classification map, and (c) reflective class legends

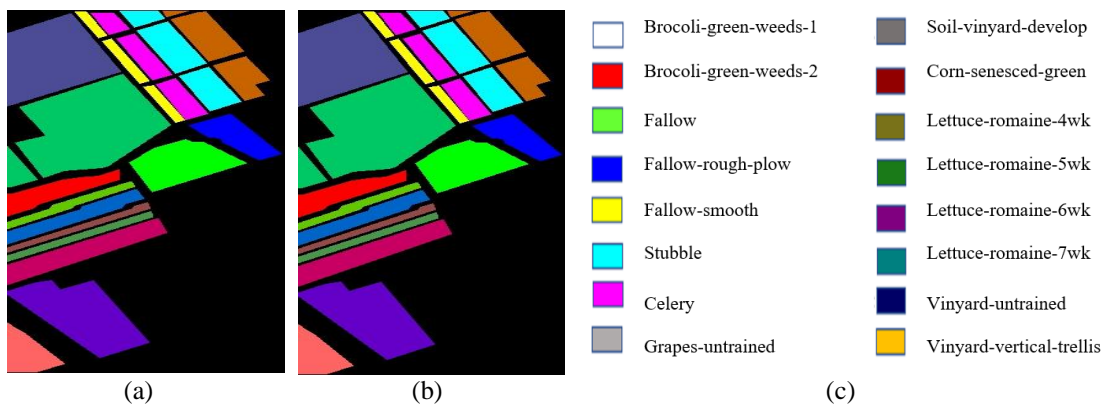


Figure 4. SA dataset; (a) ground-truth-image, (b) classification map, and (c) reflective class legends

The proposed model's classification map of Indian Institute of Space Science and Technology (IIST) dataset is compared with its respective ground truth. The classification map demonstrates that the percentage of incorrect classifications is comparatively low. Figure 5(a) represents the ground truth image of IIST dataset, Figure 5(b) represents the 3DCNN model classification map of IIST dataset, Figure 5(c) represents HybridSN model classification map of IIST dataset, and Figure 5(d) represents the proposed model classification map with reflective class legends of IIST dataset respectively.

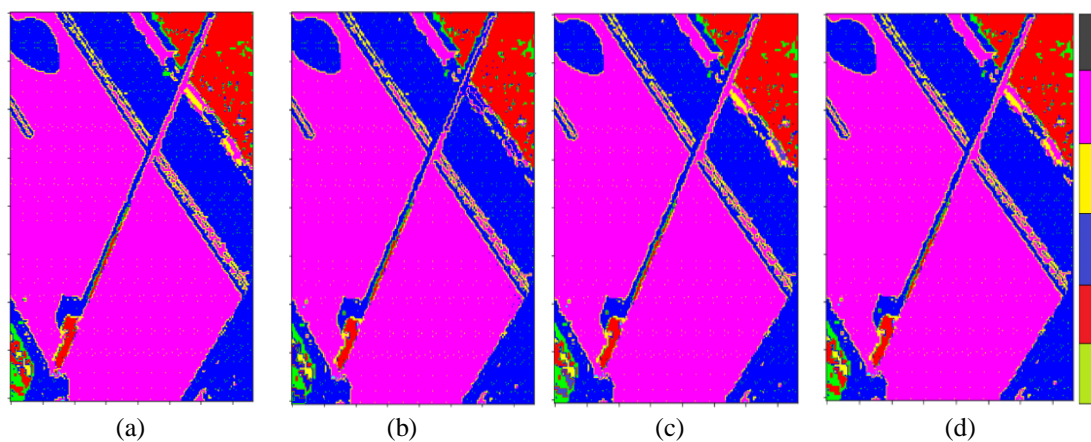


Figure 5. IIST dataset (a) ground-truth-image, (b) 3DCNN map, (c) Hybrid-SN map, and (d) proposed model classification map with reflective class legends

For Indian datasets, the proposed model's classification maps are significantly superior to 3DCNN and HybridSN approaches in terms of quality. The classification map of proposed model with Ahmedabad1 (AH1) dataset is compared to its respective ground truth. The classification map demonstrates that the percentage of incorrect classifications is comparatively low. Figure 6(a) represents the ground truth image of AH1 dataset, Figure 6(b) represents the 3DCNN model classification map of AH1 dataset, Figure 6(c) represents HybridSN model classification map of AH1 dataset, and Figure 6(d) represents the proposed model classification map with reflective class legends of AH1 dataset respectively.

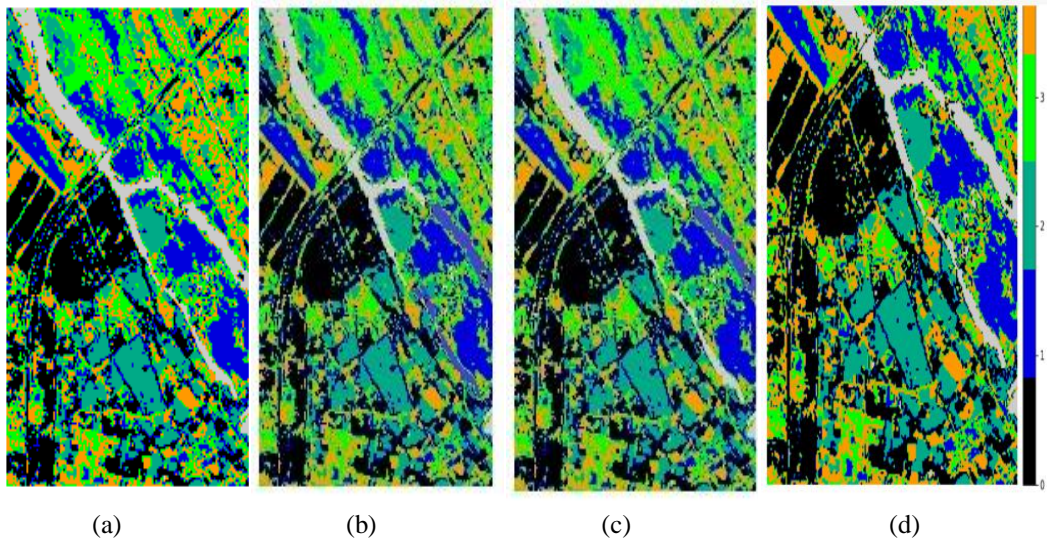


Figure 6. AH1 dataset; (a) ground-truth-image, (b) 3DCNN map, (c) hybrid-SN map, and (d) proposed model classification map with reflective class legends

The proposed model's classification map of Ahmedabad2 (AH2) dataset is compared with its respective ground truth. The classification map demonstrates that the percentage of incorrect classifications is comparatively low. Figure 7(a) represents the ground truth image of AH2 dataset, Figure 7(b) represents the 3DCNN model classification map of AH2 dataset, Figure 7(c) represents HybridSN model classification map of AH2 dataset, and Figure 7(d) represents the proposed model classification map with reflective class legends of AH2 dataset respectively.

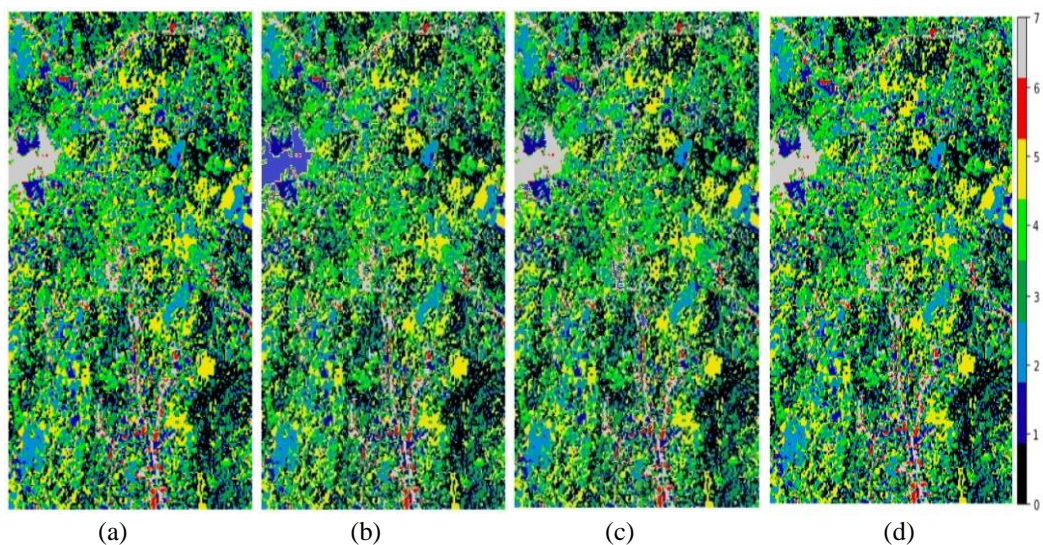


Figure 7. AH2 dataset (a) ground-truth-image, (b) 3DCNN map, (c) hybrid-SN map, and (d) proposed model classification map with reflective class legends

4. CONCLUSION

It is difficult to build a universally suitable deep learning model for hyperspectral image classification. In HSI classification, high classification accuracy can be achieved by extracting deep features from both spatial-spectral channels. However, there no theoretically valid approach to find the optimum spatial dimension to be considered. To this end, this article presented a deep CNN model, called MS-3DCNN, wherein three different multi-scale spatial-spectral patches are used to extract the deep features in both the channels. The efficiency of the proposed model is being verified through the experimental studies on three publicly available benchmark data sets and three new Indian Hyperspectral Images on which the recent methods were not tested. It is empirically proved that the classification accuracy of the proposed model is improved when compared with the remaining state-of-the-art methods used in comparative study. Further, the presented model outperformed the 3DCNN, HybridSN and LSMSC, despite having fewer training samples to work with. In the future work, the model can be further optimized to enhance the efficiency and reduce the time complexity of the model. Further, it is also to be examined to find a generalizable model to determine the optimal spatial dimension based on the data on hand in real time.




REFERENCES

- [1] M. Imani and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," *Information fusion*, vol. 59, pp. 59–83, 2020, doi: 10.1016/j.inffus.2020.01.007.
- [2] S. Prasad and J. Chanussot, *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*, Britania Raya, UK: Springer Nature, 2020, doi: 10.1007/978-3-030-38617-7.
- [3] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 120–147, 2018, doi: 10.1016/j.isprsjprs.2017.11.021.
- [4] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019, doi: 10.1016/j.isprsjprs.2019.09.006.
- [5] M. Hamouda, K. S. Ettabaa, and M. S. Bouhleb, "Smart feature extraction and classification of hyperspectral images based on convolutional neural networks," *IET Image Processing*, vol. 14, no. 10, pp. 1999–2005, 2020, doi: 10.1049/iet-ipr.2019.1282.
- [6] B. Pan, Z. Shi, and X. Xu, "Mugnet: Deep learning for hyperspectral image classification using limited samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 108–119, 2018, doi: 10.1016/j.isprsjprs.2017.11.003.
- [7] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3534–3546, June 2018, doi: 10.1109/TGRS.2018.2801387.
- [8] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep Learning-Based Classification of Hyperspectral Data," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, June 2014, doi: 10.1109/JSTARS.2014.2329330.
- [9] J. Yang, Y. Zhao, and J. C. Chan, "Learning and Transferring Deep Joint Spectral–Spatial Features for Hyperspectral Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017, doi: 10.1109/TGRS.2017.2698503.
- [10] C. Chen *et al.*, "Hyperspectral classification based on spectral–spatial convolutional neural networks," *Engineering Applications of Artificial Intelligence*, vol. 68, pp. 165–171, 2018, doi: 10.1016/j.engappai.2017.10.015.
- [11] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, "3-D Deep Learning Approach for Remote Sensing Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4420–4434, Aug. 2018, doi: 10.1109/TGRS.2018.2818945.
- [12] B. Raviteja, M. S. P. Babu, K. V. Rao, and J. Harikiran, "A New Methodology of Hierarchical Image Fusion in Framework for Hyperspectral Image Segmentation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 6, no. 1, pp. 58–65, 2017, doi: 10.11591/ijeecs.v6.i1.pp58-65.
- [13] N. He *et al.*, "Feature Extraction With Multiscale Covariance Maps for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 755–769, Feb. 2019, doi: 10.1109/TGRS.2018.2860464.
- [14] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale Dynamic Graph Convolutional Network for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3162–3177, May 2020, doi: 10.1109/TGRS.2019.2949180.
- [15] Z. Meng, L. Li, L. Jiao, Z. Feng, X. Tang, and M. Liang, "Fully dense multiscale fusion network for hyperspectral image classification," *Remote Sensing*, vol. 11, no. 22, 2019, doi: 10.3390/rs11222718.
- [16] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, Feb. 2020, doi: 10.1109/LGRS.2019.2918719.
- [17] M. Kanthi, T. H. Sarma, and C. S. Bindu, "A 3d-Deep CNN Based Feature Extraction and Hyperspectral Image Classification," *2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, 2020, pp. 229–232, doi: 10.1109/InGARSS48198.2020.9358920.
- [18] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep Learning Ensemble for Hyperspectral Image Classification," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1882–1897, June 2019, doi: 10.1109/JSTARS.2019.2915259.
- [19] X. He and Y. Chen, "Transferring CNN Ensemble for Hyperspectral Image Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 5, pp. 876–880, May 2021, doi: 10.1109/LGRS.2020.2988494.
- [20] Q. Li, B. Zheng, B. Tu, J. Wang, and C. Zhou, "Ensemble EMD-Based Spectral-Spatial Feature Extraction for Hyperspectral Image Classification," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5134–5148, 2020, doi: 10.1109/JSTARS.2020.3018710.
- [21] M. He, B. Li, and H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3904–3908, doi: 10.1109/ICIP.2017.8297014.




- [22] A. Mohan and M. Venkatesan, "HybridCNN based hyperspectral image classification using multiscale spatio-spectral features," *Infrared Physics & Technology*, vol. 108, 2020, doi: 10.1016/j.infrared.2020.103326.
- [23] K. Safari, S. Prasad, and D. Labate, "A Multiscale Deep Learning Approach for High-Resolution Hyperspectral Image Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 1, pp. 167-171, Jan. 2021, doi: 10.1109/LGRS.2020.2966987.
- [24] M. Han, R. Cong, X. Li, H. Fu, and J. Lei, "Joint spatial-spectral hyperspectral image classification based on convolutional neural network," *Pattern Recognition Letters*, vol. 130, pp. 38-45, 2020, doi: 10.1016/j.patrec.2018.10.003.
- [25] G. Sun *et al.*, "Deep fusion of localized spectral features and multi-scale spatial features for effective classification of hyperspectral images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 91, 2020, doi: 10.1016/j.jag.2020.102157.
- [26] H. Gong *et al.*, "Multiscale Information Fusion for Hyperspectral Image Classification Based on Hybrid 2D-3D CNN," *Remote Sensing*, vol. 13, no. 12, 2021, doi: 10.3390/rs13122268.
- [27] M. K. Tripathi and H. Govil, "Evaluation of aviris-ng hyperspectral images for mineral identification and mapping," *Heliyon*, vol. 5, no. 11, 2019, doi: 10.1016/j.heliyon.2019.e02931.

BIOGRAPHIES OF AUTHORS






Murali Kanthi    received the B.Tech. degree from JNTUA College of Engineering, Anantapur, Andhra Pradesh in 2007 and the M. Tech degree from JNTUA College of Engineering, Anantapur, Andhra Pradesh in 2009, where he is currently pursuing the Ph.D. degree in computer science and engineering. His research areas include Machine Learning, Hyperspectral Image Processing, Data Mining, and Deep Learning. He can be contacted at email: murali.kanthi@gmail.com



Dr. Thogarcheti Hitendra Sarma    obtained Ph. D in Machine Learning from JNT University, Anantapur, Andhra Pradesh, India in the year 2013. He is a recipient of the Teachers Associateship for Research Excellence (TARE) grant by SERB-DST Govt. of India. He has published more than 25 articles in peer-reviewed Journals and reputed international conferences like IJCNN, CEC, PReMI and others. He delivered an invited talk at FSDM -2017 in Taiwan. He is a senior member of IEEE. His research areas include Machine Learning, Hyperspectral Image Processing and Data Mining. He can be contacted at email: t.hitendrasarma@gmail.com



Dr. Chigarapalle Shoba Bindu    Ph. D in CSE from JNTUA, Anantapuramu, Andhra Pradesh. She is currently working as a Professor in the Department of CSE, JNTUA College of Engineering, Anantapuramu. Her research areas include Computer Networks, Network Security, Machine Learning, and Cloud Computing. She can be contacted at email: shobabindhu@gmail.com